

Puc Rio

Pós Graduação em Ciência de dados e Analytics

MVP Sprint 3 - Engenharia de Dados

Aluno : João Luiz Quintiliano Cassella

Objetivo :

A partir de dados sobre a Covid19 no Brasil analisar :

1. Pode-se estabelecer alguma relação entre a quantidade de casos e de mortes e o índice de Desenvolvimento Humano(IDH) ?
2. Pode-se estabelecer alguma relação entre a quantidade de casos e de mortes e a população ?
3. Pode-se estabelecer alguma relação entre a quantidade de casos e de mortes e o espectro ideológico ?

Detalhamento :

Busca pelos dados :

Utilizamos um conjunto de dados disponibilizado pelo Kaggle contendo dados sobre a pandemia de covid19 entre os anos de 2020 e 2021. Esse conjunto de dados pode ser encontrado em <https://www.kaggle.com/datasets/unanimad/corona-virus-brazil>.

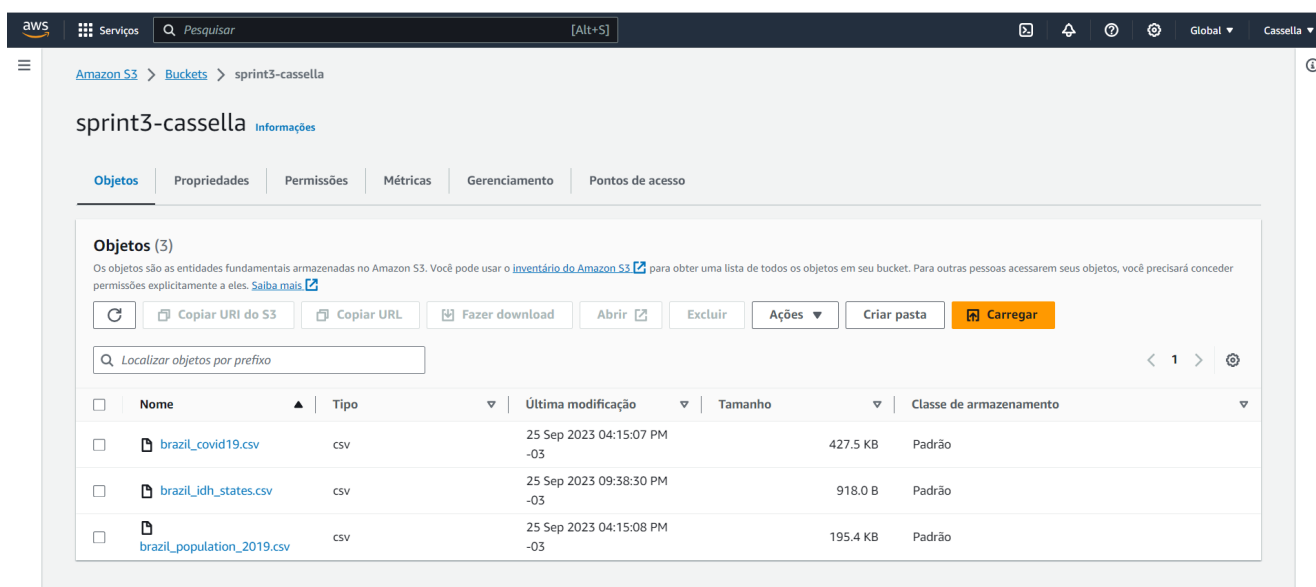
The screenshot shows the Kaggle interface for the 'Coronavirus - Brazil' dataset. On the left is a sidebar with navigation links: Home, Competitions, Datasets (selected), Models, Code, Discussions, Learn, More, Your Work, and a 'VIEWED' section listing 'Coronavirus - Brazil' and 'E-Mail classification...'. The main content area features a search bar, a user profile for 'RAPHAEL FONTES' (updated 2 years ago), and dataset statistics (409 views, 'New Notebook' button, 'Download (28 MB)' button). The dataset title 'Coronavirus - Brazil' is prominently displayed, followed by the subtitle 'Coronavirus (COVID-19) - Brazil Dataset'. Below this are tabs for 'Data Card' (selected), 'Code (83)', and 'Discussion (34)'. The 'About Dataset' section includes a request to upvote and a 'Content' section describing the dataset as 'From Novel Corona Virus 2019 Dataset'. A text box explains that the 2019 Novel Coronavirus (2019-nCoV) is a virus identified as the cause of an outbreak of respiratory illness first detected in Wuhan, China. On the right, a 'Usability' section shows a score of 10.00, a 'License' section indicating 'CC0: Public Domain', an 'Expected update frequency' of 'Daily', and a 'Tags' section with 'Health' and 'Biology' tags.

Nesse conjunto de dados, encontramos os seguintes datasets :

-
- `brazil_covid19.csv` - Quantidade de casos e mortes acumuladas por estado e data.
- `brazil_covid19_cities.csv` - Quantidade de casos e mortes acumuladas por cidade e data. (não utilizado)
- `brazil_covid19_macro.csv` - Quantidade de casos e mortes acumuladas no país e por data. (não utilizado)
- `brazil_population_2019.csv` - População por cidade em 2019.

Além desses datasets, buscamos na internet as informações do IDH por estado, além do espectro ideológico do partido dos governadores dos estados. Com essas informações, criamos o arquivo `brazil_idh_states`

Para todo o trabalho utilizamos o ambiente da Amazon (AWS) e os arquivos que iriam ser utilizados foram carregados manualmente no bucket `sprint3-cassella`.



Amazon S3 > Buckets > sprint3-cassella

sprint3-cassella [Informações](#)

[Objetos](#) | [Propriedades](#) | [Permissões](#) | [Métricas](#) | [Gerenciamento](#) | [Pontos de acesso](#)

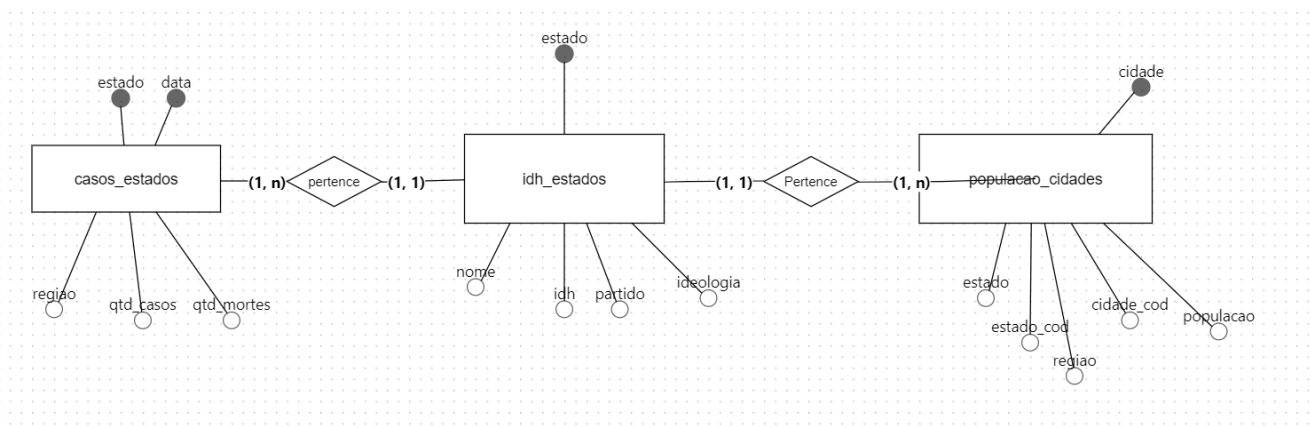
Objetos (3)

Os objetos são as entidades fundamentais armazenadas no Amazon S3. Você pode usar o [inventário do Amazon S3](#) para obter uma lista de todos os objetos em seu bucket. Para outras pessoas acessarem seus objetos, você precisará conceder permissões explicitamente a eles. [Saiba mais](#)

[Atualizar](#) [Copiar URI do S3](#) [Copiar URL](#) [Fazer download](#) [Abrir](#) [Excluir](#) [Ações](#) [Criar pasta](#) [Carregar](#)

<input type="checkbox"/>	Nome	Tipo	Última modificação	Tamanho	Classe de armazenamento
<input type="checkbox"/>	brazil_covid19.csv	csv	25 Sep 2023 04:15:07 PM -03	427.5 KB	Padrão
<input type="checkbox"/>	brazil_idh_states.csv	csv	25 Sep 2023 09:38:30 PM -03	918.0 B	Padrão
<input type="checkbox"/>	brazil_population_2019.csv	csv	25 Sep 2023 04:15:08 PM -03	195.4 KB	Padrão

Modelagem :



Obs: Como iremos utilizar o redshift, não serão implementadas as chaves primárias e nem estrangeiras.

TABELA : idh_estados						
POSIÇÃO	COLUNA	DEFAULT	NULABLE	TIPO	TAMANHO	CONTEÚDO
1	estado	NULL	SIM	character varying	256	SIGLA DO ESTADO NA FEDERAÇÃO(UF)
2	nome	NULL	SIM	character varying	256	NOME DO ESTADO
3	idh	NULL	SIM	real		ÍNDICE DE DESENVOLVIMENTO HUMANO
4	partido	NULL	SIM	character varying	256	PARTIDO A QUE O GOVERNADOR DO ESTADO É FILIADO
5	ideologia	NULL	SIM	character varying	256	ESPECTRO IDEOLÓGICO DO PARTIDO DO GOVERNADOR

TABELA : casos_estados

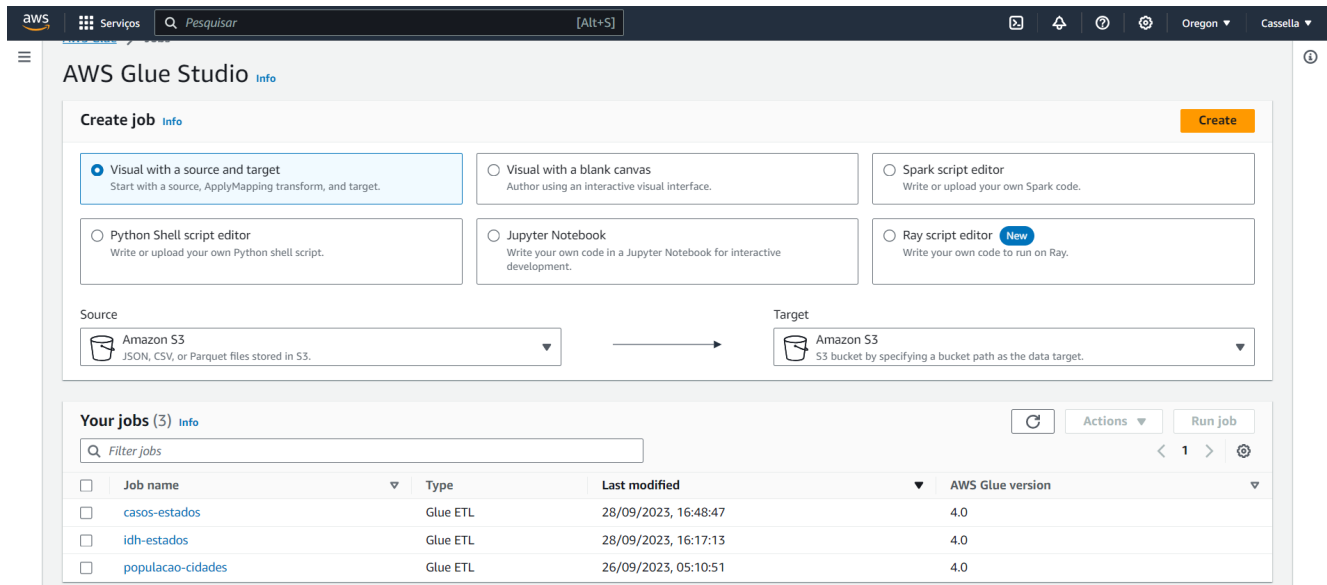
POSIÇÃO	COLUNA	DEFAULT	NULA	TIPO	TAMANHO	CONTEÚDO
1	data	NULL	SIM	date		SIGLA DO ESTADO NA FEDERAÇÃO(UF)
2	regiao	NULL	SIM	character varying	256	NOME DO ESTADO
3	estado	NULL	SIM	character varying	256	ÍNDICE DE DESENVOLVIMENTO HUMANO
4	qtd_casos	NULL	SIM	real		QUANTIDADE ACUMULADA DE CASOS ATÉ A DATA
5	qtd_mortes	NULL	SIM	real		QUANTIDADE ACUMULADA DE MORTES ATÉ A DATA

TABELA : populacao_cidades

POSIÇÃO	COLUNA	DEFAULT	NULA	TIPO	TAMANHO	CONTEÚDO
1	regiao	NULL	SIM	character varying		DESCRIÇÃO DA REGIÃO DO PAÍS
2	estado	NULL	SIM	character varying	256	NOME DO ESTADO
3	cidade	NULL	SIM	character varying		NOME DA CIDADE
4	estado_cod	NULL	SIM	bigint		CÓDIGO DO ESTADO NO IBGE
5	cidade_cod	NULL	SIM	bigint		CÓDIGO DA CIDADE NO IBGE
6	populacao	null	SIM	bigint		QUANTIDADE DE HABITANTES

Carga : Utilizamos o redshift como repositório de dados

Para o processo de ETL utilizamos o Glue para criar 3 jobs, um para cada dataset :



The screenshot shows the AWS Glue Studio interface. The top navigation bar includes the AWS logo, 'Serviços', a search bar, and regional settings for 'Oregon' and 'Cassella'. The main heading is 'AWS Glue Studio'. Below it, the 'Create job' wizard is active, showing five options: 'Visual with a source and target' (selected), 'Visual with a blank canvas', 'Spark script editor', 'Python Shell script editor', 'Jupyter Notebook', and 'Ray script editor'. The 'Source' and 'Target' dropdowns are both set to 'Amazon S3'. Below the wizard, the 'Your jobs (3)' section displays a table of existing jobs.

Job name	Type	Last modified	AWS Glue version
casos-estados	Glue ETL	28/09/2023, 16:48:47	4.0
idh-estados	Glue ETL	28/09/2023, 16:17:13	4.0
populacao-cidades	Glue ETL	26/09/2023, 05:10:51	4.0

Job populacao-cidades : Este job cria a tabela populacao_cidades

Data source properties - S3

- Name: Brasil-population-19
- S3 source type: S3 location
- S3 URL: s3://sprint3-cassella/brazil_population_2019.csv
- Recursive: ☒ Recursive
- Data format: csv
- Delimiter: Comma (,)
- Escape character: - optional
- Quote character: Double quote (")

Data target properties - Amazon Redshift

- Name: Redshift-populacao-cidades
- Node parents: Change-populacao-cidades
- Redshift access type: Direct data connection - recommended
- Redshift connection: sprint3-glue-redshift
- Connection: View properties
- Schema: public
- Table: populacao_cidades
- Handling of data and target table: TRUNCATE target table

Transform

- Name: Change-populacao-cidades
- Node parents: Brasil-population-19
- Change Schema (Apply mapping):

Source key	Target key	Data type	Drop
region	regiao	string	<input type="checkbox"/>
state	estado	string	<input type="checkbox"/>
city	cidade	string	<input type="checkbox"/>
state_code	estado_cod	bigint	<input type="checkbox"/>
city_code	cidade_cod	bigint	<input type="checkbox"/>
health_region_code			<input checked="" type="checkbox"/>
health_region			<input checked="" type="checkbox"/>
population	populacao	bigint	<input type="checkbox"/>

Na etapa 1, foram realizadas as configurações para carregar os dados do arquivo `brazil_population19.csv` que carregamos anteriormente para o bucket `sprint3-cassella`.

Na etapa 2, realizamos a transformação dos dados, convertendo os tipos dos campos "state_code" e "city_code" de string para bigint, eliminamos os campos `health_region_cod` e `health region`, além de alterar os nomes dos campos originais para criar as colunas na tabela de destino, em português.

Na etapa 3, configuramos os parâmetros para o carregamento dos dados transformados para a tabela `populacao_cidades` do redshift.

Registro da execução do job :

Services

Populacao

Alt+S

Search

Try new UI

Actions

Save

Run

populacao-cidades

Last modified on 26/09/2023, 05:10:51

Visual

Script

Job details

Runs

Data quality

Schedules

Version Control

Job runs (1/1)

Filter job runs by property

View details

Stop job run

Table View

Card View

Run status

Retries

Start time

End time

Duration

Capacity (DPUs)

Worker type

Glue version

Succeeded

0

09/26/2023 05:11:06

09/26/2023 05:13:51

2 m 9 s

10 DPUs

G.1X

4.0

09/26/2023 05:11:06

Job name

Id

Run status

Glue version

populacao-cidades

j_r_f7d93063e920d55880d5618535ccbb536a5abe4b8f8cc2b8cc62cd343b97a75

Succeeded

4.0

Retry attempt number

Start time

End time

Start-up time

Initial run

September 26, 2023 5:11:06 AM

September 26, 2023 5:13:51 AM

36 seconds

Execution time

Last modified on

Trigger name

Security configuration

2 minutes 9 seconds

September 26, 2023 5:13:51 AM

-

-

Timeout

Max capacity

Number of workers

Worker type

40 minutes

10 DPUs

10

G.1X

Execution class

Log group name

Cloudwatch logs

Performance and debugging recommendations

Standard

/aws-glue/jobs

All logs

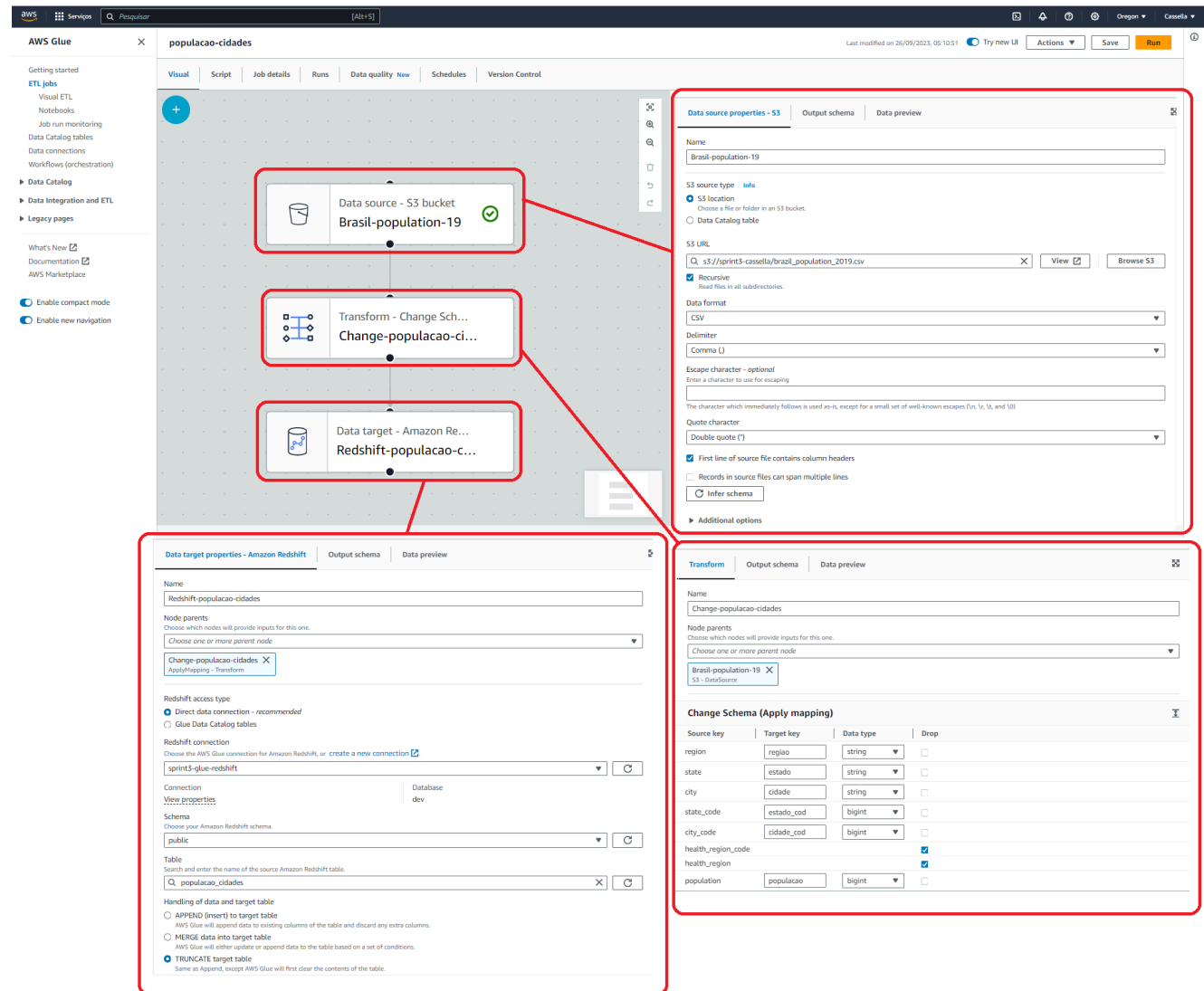
Output logs

Error logs

View in CloudWatch

Input arguments (9)

Job idh-estados : Este job cria a tabela idh_estados



Data source properties - S3

Name: Brasil-population-19

S3 source type: info

S3 location: Choose a file or folder in an S3 bucket.

S3 URL: s3://sprint3-cassella/brazil_population_2019.csv

Recursive: ☒ Read files in all subdirectories.

Data format: CSV

Delimiter: Comma (,)

Escape character - optional: Enter a character to use for escaping.

Quote character: Double quote (")

First line of source file contains column headers: ☒

Records in source files can span multiple lines: ☐

Infer schema:

Data target properties - Amazon Redshift

Name: Redshift-populacao-cidades

Node parents: Choose one or more parent node.

Change-populacao-cidades

Redshift access type: Direct data connection - recommended

Redshift connection: sprint3-glue-redshift

Connection: View properties

Schema: public

Table: populacao_cidades

Handling of data and target table: TRUNCATE target table

Transform

Name: Change-populacao-cidades

Node parents: Choose one or more parent node.

Brasil-population-19

Change Schema (Apply mapping)

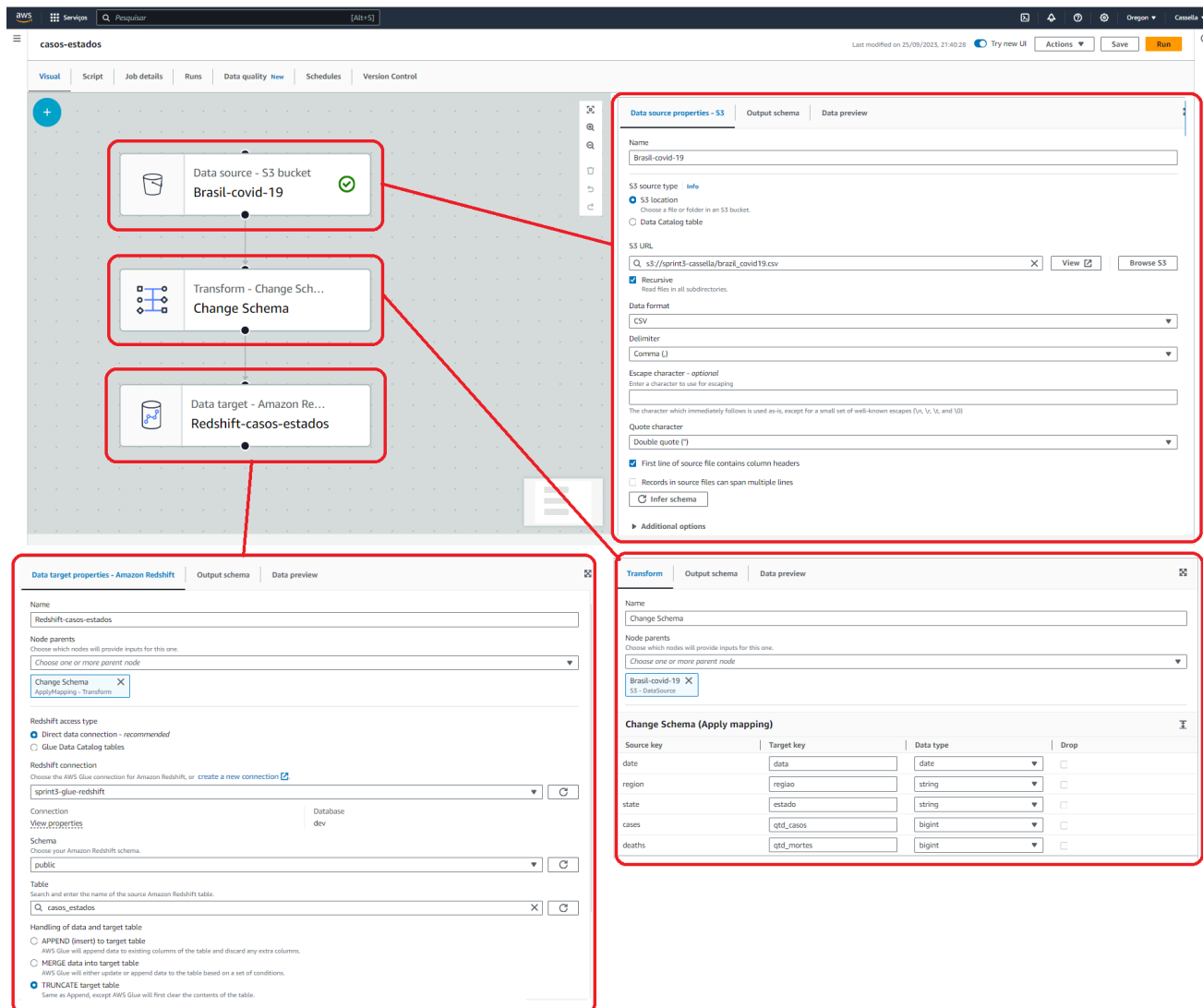
Source key	Target key	Data type	Drop
region	regiao	string	<input type="checkbox"/>
state	estado	string	<input type="checkbox"/>
city	cidade	string	<input type="checkbox"/>
state_code	estado_cod	bigint	<input type="checkbox"/>
city_code	cidade_cod	bigint	<input type="checkbox"/>
health_region_code	saude_regiao_cod	bigint	<input checked="" type="checkbox"/>
health_region	saude_regiao	bigint	<input checked="" type="checkbox"/>
population	populacao	bigint	<input type="checkbox"/>

Na etapa 1, foram realizadas as configurações para carregar os dados do arquivo `brazil_idh_states.csv` que carregamos anteriormente para o bucket `sprint3-cassella`.

Na etapa 2, realizamos a transformação dos dados, convertendo os tipos do campo "idh" de string para float, além de alterar os nomes dos campos originais para criar as colunas na tabela de destino, em português.

Na etapa 3, configuramos os parâmetros para o carregamento dos dados transformados para a tabela `idh_estados` do redshift.

Job casos-estados : Este job cria a tabela casos_estados



Data source properties - S3

Name: Brasil-covid-19

S3 source type: S3 location

S3 URL: s3://sprint3-cassella/brasil_covid19.csv

Data format: CSV

Delimiter: Comma (,)

Escape character: Double quote (")

First line of source file contains column headers: ☒

Records in source files can span multiple lines: ☐

Infer schema: ☒

Data target properties - Amazon Redshift

Name: Redshift-casos-estados

Node parents: Change Schema

Redshift access type: Direct data connection - recommended

Redshift connection: sprint3-glue-redshift

Connection: Database dev

Schema: public

Table: casos_estados

Handling of data and target table: TRUNCATE target table

Transform

Name: Change Schema

Node parents: Brasil-covid-19, S3 - DataSource

Change Schema (Apply mapping)

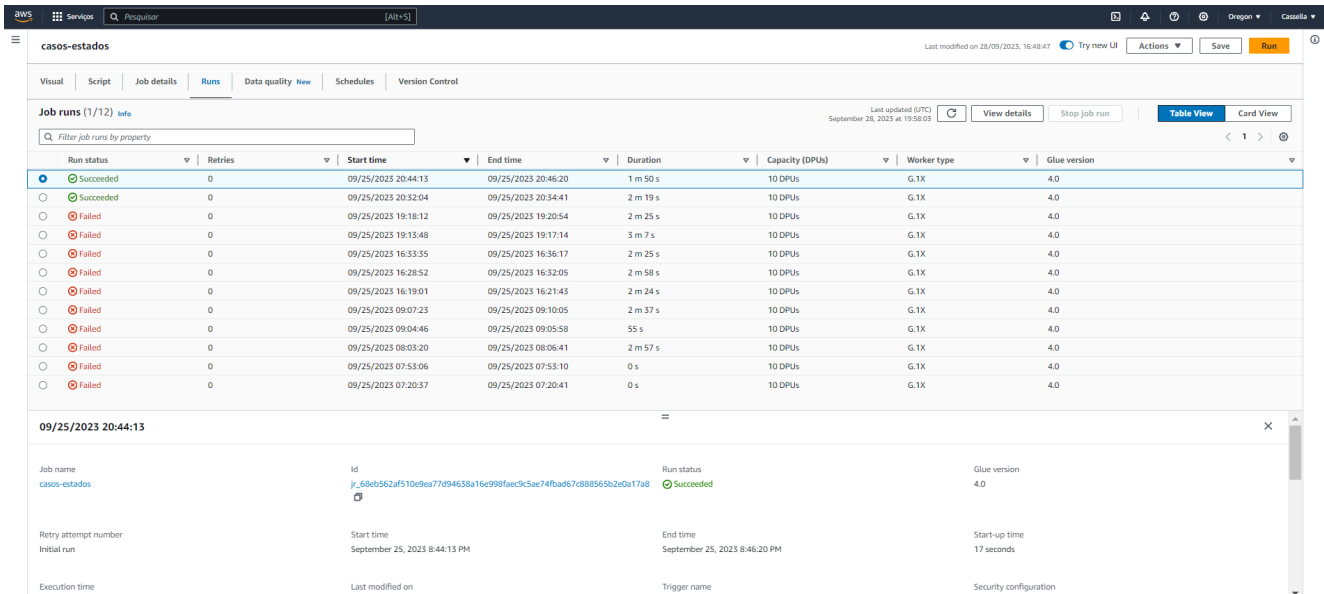
Source key	Target key	Data type	Drop
date	data	date	<input type="checkbox"/>
region	regiao	string	<input type="checkbox"/>
state	estado	string	<input type="checkbox"/>
cases	qtz_casos	bigint	<input type="checkbox"/>
deaths	qtz_mortes	bigint	<input type="checkbox"/>

Na etapa 1, foram realizadas as configurações para carregar os dados do arquivo `brasil_covid19.csv` que carregamos anteriormente para o bucket `sprint3-cassella`.

Na etapa 2, realizamos a transformação dos dados, convertendo os tipos dos campos "date" de string para date, "cases" e "deaths" de string para bigint, além de alterar os nomes dos campos originais para criar as colunas na tabela de destino, em português.

Na etapa 3, configuramos os parâmetros para o carregamento dos dados transformados para a tabela `casos_estados` do redshift.

Registro da execução do job :



Run status	Retries	Start time	End time	Duration	Capacity (DPU)	Worker type	Glue version
Succeeded	0	09/25/2023 20:44:13	09/25/2023 20:46:20	1 m 50 s	10 DPUs	G.1X	4.0
Succeeded	0	09/25/2023 20:32:04	09/25/2023 20:34:41	2 m 19 s	10 DPUs	G.1X	4.0
Failed	0	09/25/2023 19:18:12	09/25/2023 19:20:54	2 m 25 s	10 DPUs	G.1X	4.0
Failed	0	09/25/2023 19:13:48	09/25/2023 19:17:14	3 m 7 s	10 DPUs	G.1X	4.0
Failed	0	09/25/2023 16:33:35	09/25/2023 16:36:17	2 m 25 s	10 DPUs	G.1X	4.0
Failed	0	09/25/2023 16:28:52	09/25/2023 16:32:05	2 m 58 s	10 DPUs	G.1X	4.0
Failed	0	09/25/2023 16:19:01	09/25/2023 16:21:43	2 m 24 s	10 DPUs	G.1X	4.0
Failed	0	09/25/2023 09:07:23	09/25/2023 09:10:05	2 m 37 s	10 DPUs	G.1X	4.0
Failed	0	09/25/2023 09:04:46	09/25/2023 09:05:58	55 s	10 DPUs	G.1X	4.0
Failed	0	09/25/2023 08:03:20	09/25/2023 08:06:41	2 m 57 s	10 DPUs	G.1X	4.0
Failed	0	09/25/2023 07:53:06	09/25/2023 07:53:10	0 s	10 DPUs	G.1X	4.0
Failed	0	09/25/2023 07:20:37	09/25/2023 07:20:41	0 s	10 DPUs	G.1X	4.0

09/25/2023 20:44:13			
Job name	Id	Run status	Glue version
casos-estados	j_68eb562zf510e9ea77d94638a16e998fac9c5ac74fbad67c888565b2e0a17a8	Succeeded	4.0
Retry attempt number	Start time	End time	Start-up time
Initial run	September 25, 2023 8:44:13 PM	September 25, 2023 8:46:20 PM	17 seconds
Execution time	Last modified on	Trigger name	Security configuration

Observação sobre o processo de ETL :

Tivemos muita dificuldade para realizar o processo de ETL pois enfrentamos muitos problemas com a conversão de tipagem de dados. Inicialmente estávamos criando a tabela no redshift e associando na etapa 2 do job. Ao executar, recebíamos um erro genérico que depois identificamos como sendo de conversão de tipagem de dados. A solução encontrada foi de não criar a tabela, só informando seu nome na etapa 2 do job e deixando o glue criar a tabela no redshift. Para isso tomamos o cuidado de configurar o parâmetro “handling data and target table” na etapa 3 dos jobs, com a opção TRUNCATE. Testamos a opção DROP e também funcionou.

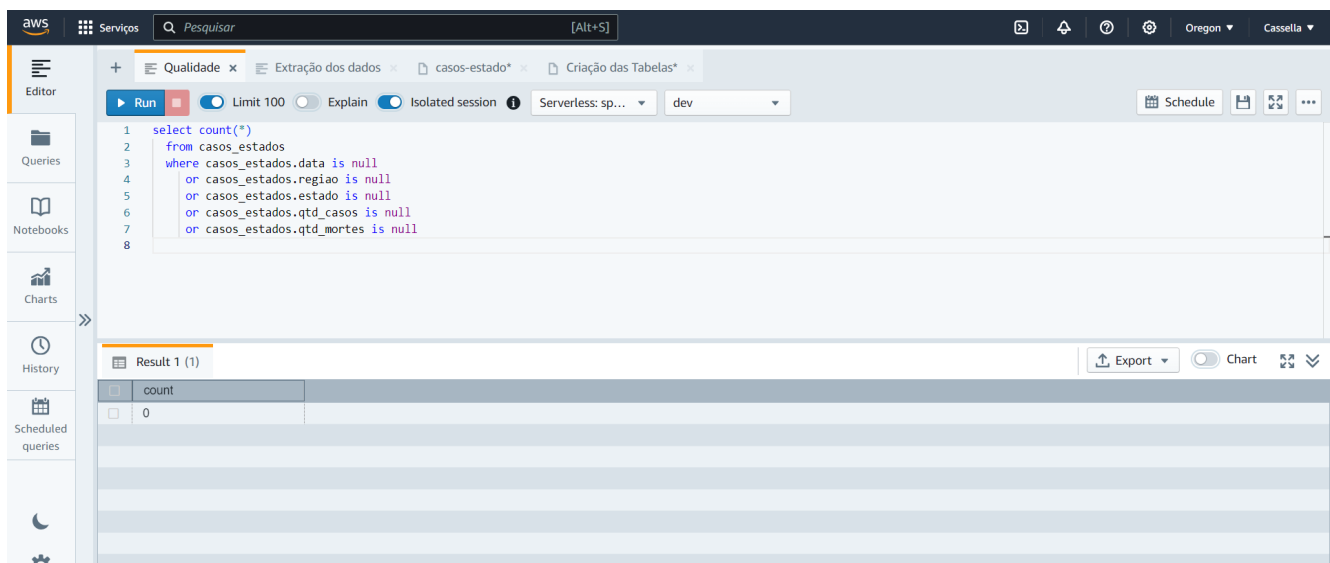
Análise :

Qualidade dos dados :

Para avaliar a qualidade dos dados, utilizamos consultas sql às tabelas do redshift .

- Tabela casos_estados :

A tabela não contém nenhum registro com informações nulas



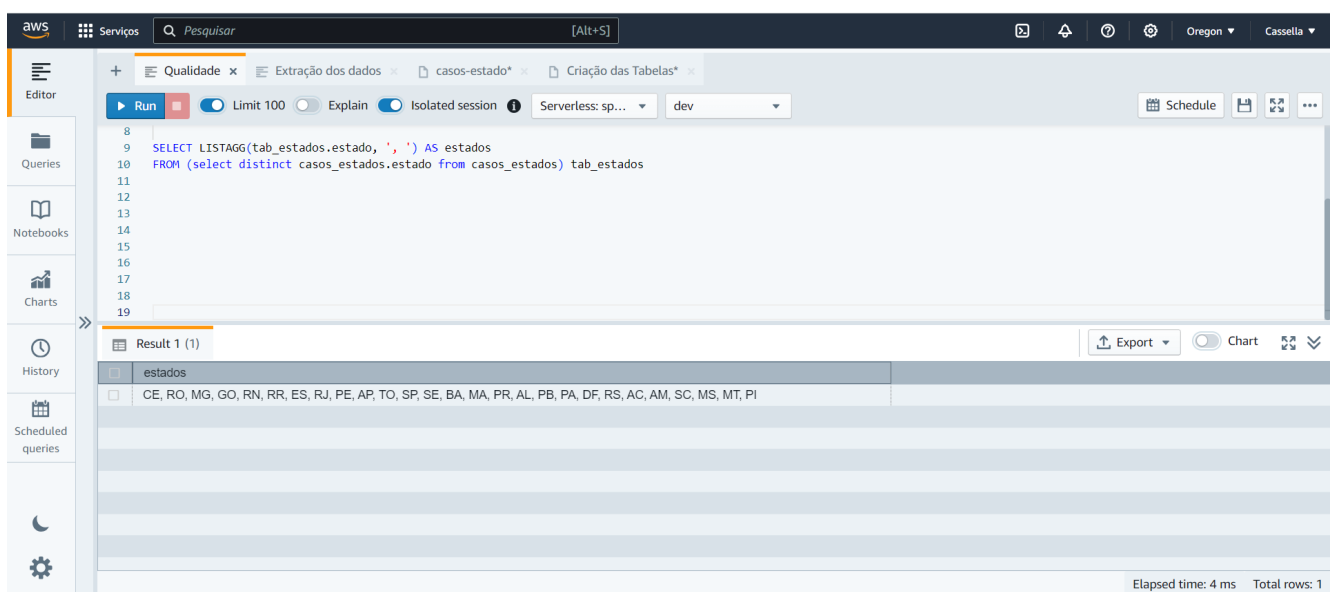
The screenshot shows the AWS Redshift console interface. The SQL editor contains the following query:

```
1 select count(*)
2 from casos_estados
3 where casos_estados.data is null
4    or casos_estados.regiao is null
5    or casos_estados.estado is null
6    or casos_estados.qtd_casos is null
7    or casos_estados.qtd_mortes is null
8
```

The query is executed, and the results are displayed in a table with one row and one column:

count
0

Todos os estados constantes nos registros da tabela são válidos



The screenshot shows the AWS Redshift console interface. The SQL editor contains the following query:

```
8
9 SELECT LISTAGG(tab_estados.estado, ', ' ) AS estados
10 FROM (select distinct casos_estados.estado from casos_estados) tab_estados
11
12
13
14
15
16
17
18
19
```

The query is executed, and the results are displayed in a table with one row and one column:

estados
CE, RO, MG, GO, RN, RR, ES, RJ, PE, AP, TO, SP, SE, BA, MA, PR, AL, PB, PA, DF, RS, AC, AM, SC, MS, MT, PI

At the bottom right of the console, it shows "Elapsed time: 4 ms" and "Total rows: 1".

- Tabela idh_estados :

Na tabela não existem registros com colunas sem informação (null)

The screenshot shows the AWS Glue console interface. The query editor contains the following SQL code:

```
select count(*)
from idh_estados
where idh_estados.estado is null
or idh_estados.nome is null
or idh_estados.idh is null
or idh_estados.partido is null
or idh_estados.ideologia is null
```

The query is executed, and the results are displayed in a table with the following structure:

count
0

The status bar at the bottom indicates "Elapsed time: 2292 ms" and "Total rows: 1".

Todos os estados constantes nos registros da tabela são válidos

The screenshot shows the AWS Glue console interface. The query editor contains the following SQL code:

```
SELECT LISTAGG(tab_estados.estado, ', ') AS estados
FROM (select distinct idh_estados.estado from idh_estados) tab_estados
```

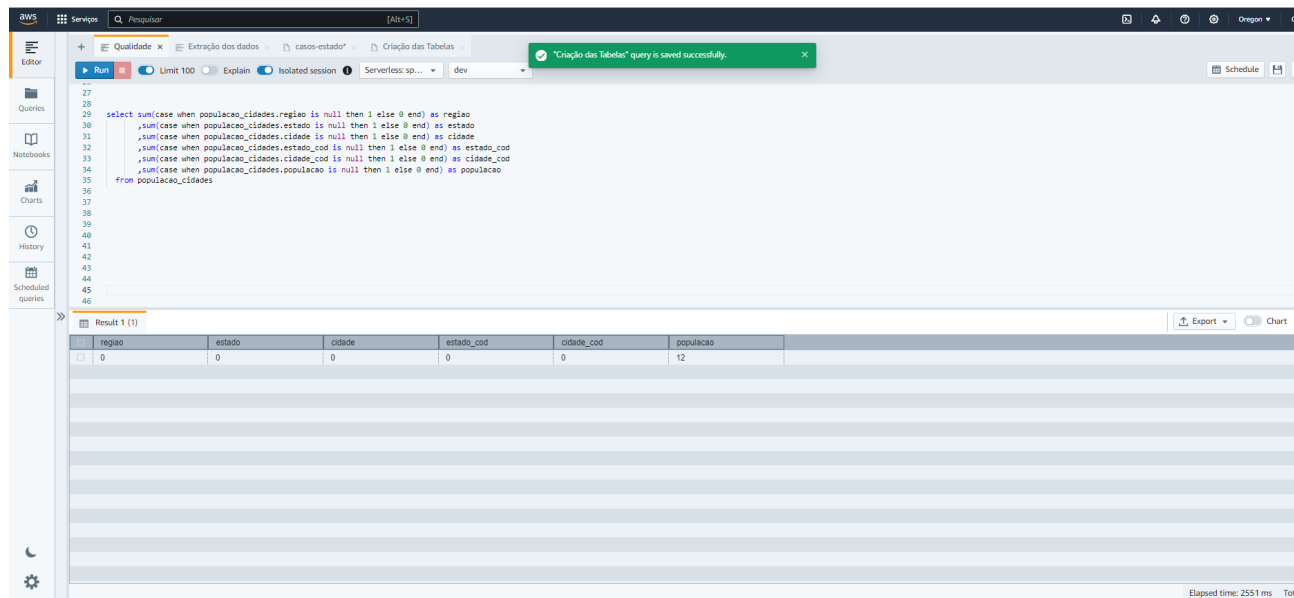
The query is executed, and the results are displayed in a table with the following structure:

estados
DF, SP, SC, MG, RS, ES, PR, RJ, MS, GO, MT, CE, TO, RN, PE, AC, SE, AM, RO, RR, PB, BA, PI, PA, AP, AL, MA

The status bar at the bottom indicates "Elapsed time: 2297 ms" and "Total rows: 1".

- Tabela populacao_cidades :

Na tabela foram encontrados 12 registros com a coluna populacao = null.
Essas cidades serão excluídas da análise.



The screenshot shows the AWS Glue console interface. A SQL query is being executed, and the results are displayed in a table. The query is as follows:

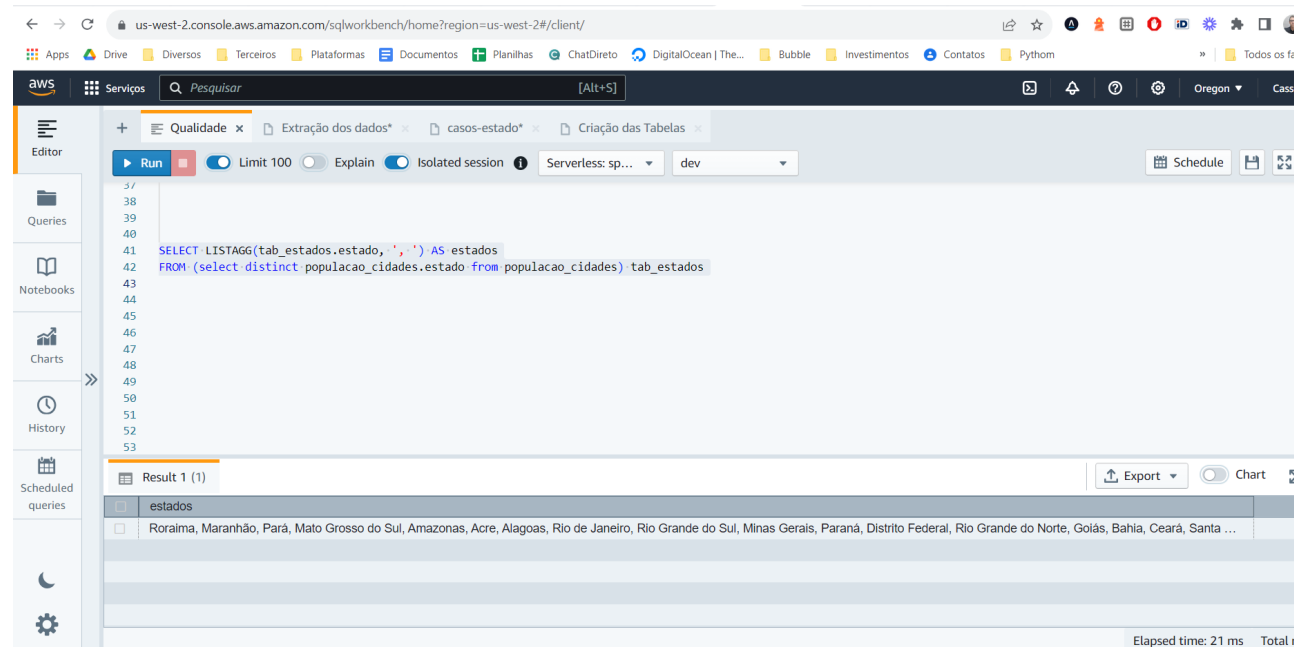
```
select sum(case when populacao_cidades.regiao is null then 1 else 0 end) as regiao,
sum(case when populacao_cidades.estado is null then 1 else 0 end) as estado,
sum(case when populacao_cidades.cidade is null then 1 else 0 end) as cidade,
sum(case when populacao_cidades.estado_cod is null then 1 else 0 end) as estado_cod,
sum(case when populacao_cidades.cidade_cod is null then 1 else 0 end) as cidade_cod,
sum(case when populacao_cidades.populacao is null then 1 else 0 end) as populacao
from populacao_cidades
```

The results table shows the following data:

regiao	estado	cidade	estado_cod	cidade_cod	populacao
0	0	0	0	0	12

The interface also shows a notification: "Criação das Tabelas" query is saved successfully.

Todos os estados constantes nos registros da tabela são válidos



The screenshot shows the AWS Glue console interface. A SQL query is being executed, and the results are displayed in a table. The query is as follows:

```
SELECT LISTAGG(tab_estados.estado, ',' AS estados
FROM (select distinct populacao_cidades.estado from populacao_cidades) tab_estados
```

The results table shows the following data:

estados
Roraima, Maranhão, Pará, Mato Grosso do Sul, Amazonas, Acre, Alagoas, Rio de Janeiro, Rio Grande do Sul, Minas Gerais, Paraná, Distrito Federal, Rio Grande do Norte, Goiás, Bahia, Ceará, Santa ...

The interface also shows a notification: "Criação das Tabelas" query is saved successfully.

Solução das Questões

A etapa de análise foi realizada através de scripts sql executados no ambiente de consulta de dados do redshift. Foram executados 2 scripts :

Script 1 : Realizamos uma consulta para retornar o estado, idh, população do estado(em milhões de habitantes), o % de casos em função da população e o % de mortes em função da quantidade de casos. Numa primeira execução ordenamos o resultado pela coluna idh e numa segunda execução, pela coluna população.

```

1 select idh_estados.estado
2     ,replace((idh_estados.idh*100)::text,',','.') as idh
3     ,replace((v_populacao.populacao::numeric/1000000)::text,',','.') AS populacao
4     ,replace((v_dados.total_casos*100/v_populacao.populacao)::text,',','.') as casos_populacao
5     ,replace((v_dados.total_mortes*100/v_dados.total_casos)::text,',','.') as mortes_casos
6 from idh_estados
7 join (
8     select populacao_cidades.estado
9         ,sum(populacao_cidades.populacao) as populacao
10        from dev.public.populacao_cidades
11        where dev.public.populacao_cidades.populacao is not null
12        group by populacao_cidades.estado
13        ) v_populacao on v_populacao.estado = idh_estados.nome
14 join (
15     select casos_estados.estado
16         ,max(casos_estados.qtd_casos) total_casos
17         ,max(casos_estados.qtd_mortes) total_mortes
18        from dev.public.casos_estados
19        group by casos_estados.estado
20        ) v_dados on v_dados.estado = idh_estados.estado
21 order by idh_estados.idh
22         --v_populacao.populacao
23

```

Script 2 : Realizamos uma consulta para retornar a ideologia (espectro ideológico do partido político a que pertencia o governador do estado), o % de casos em função da população e o % de mortes em função da quantidade de casos, agrupando as informações pela coluna ideologia.

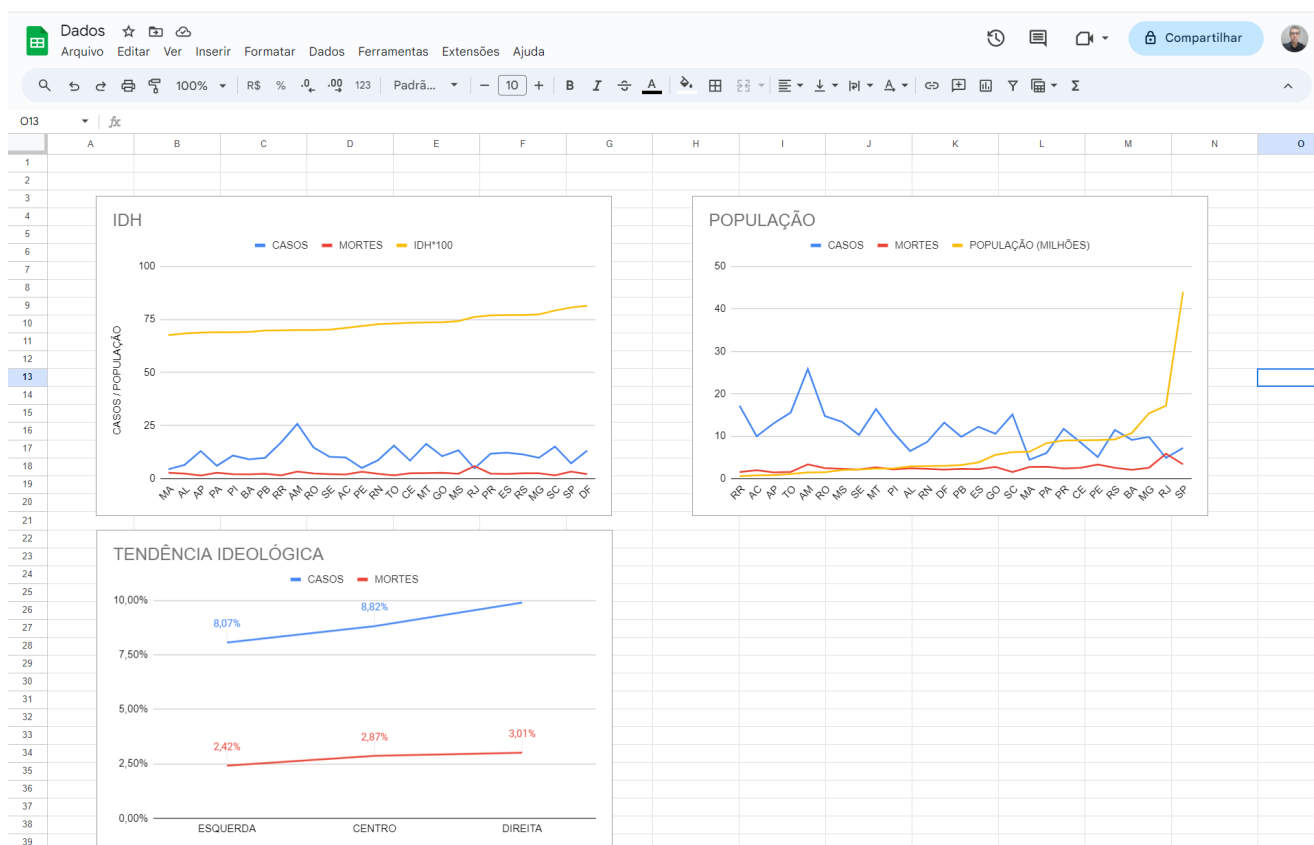
```

26
27 select idh_estados.ideologia as ideologia
28     ,replace(to_char((sum(v_dados.total_casos)/sum(v_populacao.populacao)), '009999'),' ','.') as casos_populacao
29     ,replace(to_char((sum(v_dados.total_mortes)/sum(v_dados.total_casos)), '009999'),' ','.') as mortes_casos
30 from idh_estados
31 join (
32     select populacao_cidades.estado
33         ,sum(populacao_cidades.populacao) as populacao
34        from dev.public.populacao_cidades
35        where dev.public.populacao_cidades.populacao is not null
36        group by populacao_cidades.estado
37        ) v_populacao on v_populacao.estado = idh_estados.nome
38 join (
39     select casos_estados.estado
40         ,max(casos_estados.qtd_casos) total_casos
41         ,max(casos_estados.qtd_mortes) total_mortes
42        from dev.public.casos_estados
43        group by casos_estados.estado
44        ) v_dados on v_dados.estado = idh_estados.estado
45 group by idh_estados.ideologia
46
47
48
49
50
51

```

Após a execução dos scripts, transferimos as séries de dados para uma planilha do Google e geramos os gráficos para análise.

Dados																			
Arquivo Editar Ver Inserir Formatar Dados Ferramentas Extensões Ajuda																			
75% 123 Padrão 10 B I A																			
S6																			
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	ESTADO	IDH	POPULAÇÃO	CASOS	MORTES		ESTADO	IDH	POPULAÇÃO	CASOS	MORTES		IDEOLOGIA	CASOS	MORTES				
2	MA	67.600	6	4.460786483	2.777338103		RR	0.699	0.59381	17.18038231	1.570295729		ESQUERDA	8,07%	2,42%				
3	AL	68.400	3	6.52233343	2.33107939		AC	0.710	0.81438	9.96603018	2.00272776		CENTRO	8,82%	2,87%				
4	AP	68.800	1	13.05415079	1.500065013		AP	0.688	0.84573	13.05415079	1.500065013		DIREITA	9,90%	3,01%				
5	PA	69.000	8	6.066815053	2.799302977		TO	0.731	1.11059	15.58712036	1.603613908								
6	PI	69.000	2	10.91604455	2.172568774		AM	0.700	1.48076	25.83050426	3.377892582								
7	BA	69.100	11	9.144645454	2.076887962		RO	0.700	1.52671	14.70044223	2.489915842								
8	PB	69.800	3	9.955400911	2.341454794		MS	0.742	2.06397	13.41478482	2.333527162								
9	RR	69.900	1	17.18038231	1.570295729		SE	0.702	2.18127	10.32689747	2.176180984								
10	AM	70.000	1	25.83050426	3.377892582		MT	0.736	2.36510	18.41723506	2.681788813								
11	RO	70.000	2	14.70044223	2.489915842		PI	0.690	2.42960	10.91604455	2.172568774								
12	SE	70.200	2	10.32689747	2.176180984		AL	0.684	2.89709	6.52233343	2.438107939								
13	AC	71.000	1	9.96603018	2.00272776		RN	0.728	2.96620	8.678502449	2.322256839								
14	PE	71.800	9	5.086515807	3.323110596		DF	0.814	3.01527	13.21441411	2.124733342								
15	RN	72.800	3	8.678502449	2.322256839		PB	0.698	3.21221	9.865400911	2.341454794								
16	TO	73.100	1	15.58712036	1.603613908		ES	0.771	3.84155	12.23516231	2.236500574								
17	CE	73.400	9	8.521972591	2.571633686		GO	0.737	5.62103	10.59108364	2.781491883								
18	MT	73.600	2	16.41723506	2.681788813		SC	0.762	6.26052	15.13947955	1.563815533								
19	GO	73.700	6	10.59108364	2.781491883		MA	0.676	6.37334	4.460786483	2.777338103								
20	MS	74.200	2	13.41478482	2.333527162		PA	0.690	8.38081	6.066815053	2.799302977								
21	RJ	76.200	17	4.889573297	5.891276413		PR	0.769	9.01823	11.76154301	2.404676986								
22	PR	76.900	9	11.76154301	2.404676986		CE	0.734	9.04054	8.521972591	2.571933686								
23	ES	77.100	4	12.23516231	2.236500574		PE	0.719	9.09007	5.086515807	3.323110596								
24	RS	77.100	9	11.46729262	2.586722516		RS	0.771	9.24359	11.46729262	2.586722516								
25	MG	77.400	15	9.857736923	2.577752687		BA	0.691	10.75508	9.144645454	2.076887962								
26	SC	79.200	6	15.13947955	1.563815533		MG	0.774	15.38164	9.857736923	2.577752687								
27	SP	80.600	44	7.24415347	3.37548481		RJ	0.762	17.18923	4.889573297	5.891276413								
28	DF	81.400	3	13.21441411	2.124733342		SP	0.806	44.00935	7.24415347	3.37548481								



Conclusões :

Questão 1 : Pode-se estabelecer alguma relação entre a quantidade de casos e de mortes e o índice de Desenvolvimento Humano(IDH) ?

Análise : Olhando o gráfico IDH acima, observamos que no eixo x temos os estados dispostos em ordem crescente do IDH e podemos verificar que embora o IDH cresça ao longo do eixo, a quantidade de casos e de mortes tem um comportamento aleatório, variando para cima e para baixo sem uma relação direta com o aumento do IDH. Podemos concluir portanto que não podemos estabelecer uma relação dos casos e mortes de covid com o IDH;

Questão 2 : Pode-se estabelecer alguma relação entre a quantidade de casos e de mortes e a população ?

Análise : Olhando o gráfico POPULAÇÃO acima, observamos que no eixo x temos os estados dispostos em ordem crescente da sua população e podemos verificar que embora a população cresça ao longo do eixo, a quantidade de casos e de mortes tem um comportamento aleatório, variando para cima e para baixo sem uma relação direta com o aumento da população. Podemos concluir portanto que não podemos estabelecer uma relação dos casos e mortes de covid com a população;

Questão 3 : Pode-se estabelecer alguma relação entre a quantidade de casos e de mortes e o espectro ideológico?

Análise : Olhando o gráfico TENDÊNCIA IDEOLÓGICA acima, observamos que no eixo x podemos verificar os espectros políticos ESQUERDA, CENTRO e DIREITA. Existe uma pequena variação nos casos de covid em torno de 1% entre o menor valor(ESQUERDA) e o maior (DIREITA) e em torno de 0,2% nas mortes. Consideramos que apesar dessa variação existir, não é significativa o bastante para estabelecer uma relação com o espectro político e mais uma vez chegamos à conclusão de que não podemos estabelecer uma relação dos casos e mortes de covid com o espectro político;